DOCUMENT RESUME

ED 308 708 FL 018 071

AUTHOR Pinkham, Jessie
TITLE A Grammar of English for a Machine Translation System.
PUB DATE Oct 86
NOTE 10p.; In: Proceedings of the Annual Eastern States Conference on Linguistics (3rd, Pittsburgh, PA, October 10-11, 1986). The complete Proceedings are available from the Dept. of Linguistics, Ohio State University, Columbus.
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *English; *Grammar; *Language Processing; *Machine Translation; Phrase Structure; Structural Analysis (Linguistics)
IDENTIFIERS *Left Modifier Grouping; Parsing

ABSTRACT

The development of a new grammar of English for machine translation systems at Weidner Communications Corporation is described. Although the project began with the intention of simply modifying the grammar rules already in use in the translation system, the reorganization of the grammar was deemed necessary for providing a principled manner of increasing coverage and for improving the efficiency of the grammar read by the parser. The most innovative aspect of the grammar is the approach to the internal structure of phrasal constituents. This approach, referred to as Left-Modifier Grouping, groups left-hand modifiers of most lexical heads with the head as a constituent to which right-hand modifiers are then attached. Left Modifier Grouping is based on a linguistic generalization about the phrase structure of English by which nodes to the left of all lexical heads except "V" are not themselves right-branching. A consequence of Left Modifier Grouping is that a practical grammar of English for parsing can be broken down into two types of rules: (1) those that are not recursive, thus applicable only once; and (2) those that are recursive, allowing reapplication. A summary of the two blocks of rules is presented. (DJD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

BEST COPY AVAILABLE

ED308708

A Grammar of English for a Machine Translation System

by

Jessie Pinkham

from

Proceedings of the Third

EASTERN STATES CONFERENCE ON LINGUISTICS

The University of Pittsburgh

Pittsburgh, Pennsylvania

October 10-11, 1986

Fred Marshall
Editor
The University of Pittsburgh

Ann Miller
Associate Editor
The Ohio State University

Zheng-sheng Zhang
Associate Editor
The Ohio State University

FL018071

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Joseph, B D

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

A GRAMMAR OF ENGLISH FOR A MACHINE TRANSLATION SYSTEM*

Jessie Pinkham
WCC

The grammar outlined in this paper was developed in the context of a machine translation system at Weidner Communications Corporation (Northbrook, IL). We began this project with the intention of simply modifying the grammar rules which were already in use in the translation system. However, it became necessary to modify the organization of the grammar, both to provide a principled manner for increasing coverage, and to improve the efficiency with which the grammar was read by the parser.

The main innovative aspect of the grammar is the approach to the internal structure of phrasal constituents. This approach, referred to as Left Modifier Grouping, groups left-hand modifiers of most lexical heads with the head as a constituent to which right-hand modifiers are then attached. Left Modifier Grouping is based on a linguistic generalization about the phrase structure of English, by which nodes to the left of all lexical heads except V are not themselves right-branching. The linguistic validity of this generalization is argued in section 2.

A consequence of Left Modifier Grouping is that a practical grammar of English for parsing can be broken down into two types of rules: those that are not recursive, and thus can be applied only once, and those that are recursive, and must be allowed to reapply. This strategy assumes that the rules of the grammar are matched on the input by the parser. A summary of the two blocks of rules created is presented in section 3.

1. Context

This grammar of English is part of the analysis component of an existing machine translation system which translates from English into French, Spanish, German, Italian, Portuguese and Arabic. The translation system is automatic in that it takes input and produces output without any human intervention, although the output requires human post-editing. The texts input to the system by users are typically technical material.

The analysis component uses a bottom-up parser, which is deterministic in the sense that it produces the first parse found. When input reaches the parser, it has been morphologically analyzed, and all lexical ambiguities have been resolved. The grammar consists of ordered rules, which may be either context-free or context-sensitive. The output of the parser is a syntactic tree annotated with semantic, syntactic and functional features.

The analysis module satisfies two minimal requirements dictated by its use in a non-interactive machine translation system. First, it is robust enough to handle both grammatical and ungrammatical input, and produces partial parses when a complete

parse cannot be obtained. Second, it permits identification of arguments of verbs, adjectives and nouns, which is crucial for correct lexical transfer (for example, the translation of the verb "work" varies in French depending on the animacy of the subject).

2. Left Modifier Grouping

Emonds (1976) examines constraints on phrase structure configurations. He notes that subjects of verbs are unusual in exhibiting totally free recursion, and goes on to observe that this is not usually the case for left-hand modifiers in English [1]:

> "By contrast, such free surface recursion is not exhibited in English by possessive noun phrases (the King of England's hat), prenominal adjective phrases (three tiring miles long), intensifying adverbial adjective phrases (so magnificently attired, quite simply boring), preverbal adverbial adjective phrases (politely put it away, very quickly finished his work), or prepositional measure phrases (put it three feet beyond the table). I have not investigated in detail how this restriction on recursion in surface structure can best be stated, but it certainly seems to exist."

A constraint on free recursion exists, as formulated in (1), although it can be violated under certain discourse conditions, and in a few exceptional environments. The constraint proves useful in describing written English, particularly technical material, and allows us to develop a PS grammar of English which provides empirically realistic coverage.

1) Constraint on free recursion
In the configuration $[..A..H..]_{H2}$ where H is a lexical head, and A is the modifier of this head, the node A will not exhibit free recursion, i.e., it will not dominate S, PP, or VP.

The following examples illustrate the coverage of the constraint:

2) *The happy { to work / that you like him / with his job } man ...

3) *[The man who is hanging up his hat]'s wife ...
4) *It is more than 5 pounds heavy.
5) *He was going much more than 50 mph fast.

6) The meeting that she was going to was rescheduled.

The constraint operates in examples (2-5), for the lexical heads N, Adj, Adv. It is assumed to work for Q and Prep as well, but plausible examples to show its effect are impossible to construct. In example (6), where the lexical head is V, it does not hold. The exclusion of the head V from this constraint demonstrates a significant difference between modifiers of nouns, adjectives or adverbs and subjects of verbs which is difficult to account for under any version of X bar theory which treats them in the same way. The Left Modifier Grouping perspective on phrase structure assumes that a subject noun phrase is not the left-hand modifier of a verb, and that S is not the maximal projection of V. The exclusion of V from the class of lexical heads for which the constraint applies therefore follows in a straightforward way.

There are instances where the constraint does not seem to hold:

7) ?An easy to open jar...

8) The man who lives next door's oldest son...
The King of England's hat...

9) He is more than 6 feet tall.

In (7), the left-hand modifier of the noun is an adjective with a VP complement. The interesting fact, observed in Nanni (1980), is that the VP complement allowed in this position is syntactically restricted. While the simple VP "to open" is permitted, the more complex VPs "for us to open" or "to open with one hand" are not. Furthermore, the adjectives which allow the syntactic pattern in (7) are only "easy/tough" adjectives. The important point for the discussion here is that the phrase structure rules which will be required to produce the adjective phrase in examples such as (7) do not need to be recursive. The pattern "Adj to V(inf)" can be viewed as a case of special frozen adjective phrase, applying only in the cases of "easy/tough" adjectives [2].

The contrast between (9) and (4-5) is also very interesting. The ability to have "more than X" as a left-hand modifier of an adjective coincides with the possibility of X alone as the left-hand modifier of the adjective.

10) *He is 160 pounds heavy.
*He is going 50 mph fast.

11) He is 6 feet tall.

Examples (10), like (4-5), are not permitted. Sentence (11), like (9), is allowed. Further examples of this correlation are given below. The (a) sentences contain the adjectives modified by a non-comparative measure, and the (b) sentences contain the adjectives modified by the comparative form.

12a) The tower is 40 feet high.
b) The tower is more than 40 feet high.

13a) The man is 60 years old.
b) The man is more than 60 years old.

14a) The table is 3 feet long.
b) The table is more than 3 feet long.

15a) *The temperature is 20 degrees high.
b) *The temperature is more than 20 degrees high.

16a) *Her mother is 5 feet short.
b) *Her mother is less than 5 feet short.

The original contrast between the unacceptability of the forms with "heavy" (example (3)) and the grammaticality of the forms with "tall" (example (11)) is clearly due to the fact that "heavy" does not allow a left-hand measure modifier, while "tall" does. Given this, the internal structure of the phrase "more than 6 feet tall" is crucial to determining whether (11) violates the constraint on free recursion. We assume, as in Pinkham (1985), that the internal structure of the comparative "more than Q N Adj" is "[[more than Q] N] Adj]" and that "more than Q" is a parallel comparative, i.e., similar to a coordinate structure. [more than Q] can be formed by a "coordinate" quantifier rule, not a quantifier phrase rules. Since the modifier of the adjective does not contain a PP, these sentences are not a problem for the constraint on free recursion.

The examples in (8) involving possessive noun phrases present a more difficult problem. It is possible to have modified noun phrases such as "the King of England" as possessive noun phrases. Here, as in the "easy to open" example, however, the PP is not completely free. Note the contrast between the (a) examples and the (b) examples:

8a) The King of England's hat...
b) ?*The King of any small island in the Northern hemisphere's major concern...

17a) The oldest bridge's renovation ...
b) ?*The oldest bridge over the Ohio river's renovation

18a) A movie director's attitude..
b) ?*A movie director of any standing's attitude..

We consider that the "King of England" is a complex noun which does not dominate a (freely recursive) PP. Other examples are "Governor of the State of Illinois", and the "President of the

United States. A possible structure is [N1 of N1], which would apply to proper nouns [3].

Similarly, relative clauses cannot appear freely in possessives:

3) ?*The man who is hanging up his hat's wife ...
8) The man who lives next door's oldest son ...
19) ?*The man who is picking up his wife's son ...[4]

The judgements given on the ungrammatical sentences depend on the discourse context that one is considering. They can be acceptable under what several speakers call the "hyphen-reading", i.e., one where the possessive seems to be grouped as a hyphenated string. The hyphen-reading requires an intonational break after the end of the possessive, and becomes much more plausible in the right discourse context. Imagine this exchange taking place at a meeting of the United Nations:

20) A: "The King of Ponape's most pressing concern is the bad influence of tourism."
B: "Well, the King of any small island's concern would be the same."

The reply can be elaborated, but acceptability does go down:

21) B: "Well, the King of any small island which is being overrun by tourists 's concern would be the same."

Here again, a speaker would rely on special intonation to make clear that the possessive formed a unit.

We cannot provide a complete account of the conditions which allow PPs and relative clauses in possessives. These cases therefore remain a problem for the constraint on free recursion as stated in (1). For our purposes, however, we can note that these violations are very rare in written English, and non-existent in technical material, and that a grammar based on the constraint in (1) remains empirically correct for written English.

The Left Modifier Grouping approach considers the nodes X2 as the nodes which can freely recurse. Only X2 is considered a phrase. Nodes which do not exhibit free recursion are X1. The approach can be summarized as follows:

22) Left Modifier Grouping
A head X combines with its left-hand modifiers to form X1
X1 combines with right-hand modifiers to form X2
X1 is not considered a phrase

An example of the application of Left Modifier Grouping is the noun phrase in (23):

23) [[the [[very] tall] player] [who was hired]]
Adv1 Adj1 N1 N2

This can be contrasted to a conventional approach, where Right Modifier Grouping applies:

24) [[the [[very] tall] [player [who was hired]]]
Adv2 Adj2 N1 N2

3. Outline of the grammar

We have constructed a grammar for our application based on the Left Modifier Approach, which is divided into two blocks. The first block builds all X1 constituents, which as we stated above, do not exhibit free recursion. The rules are ordered, and are applied only once. This block is referred to as the non-cyclic rules:

25) Non-cyclic rules: ordered tables of rules covering
Hyphenated structures
Compound nouns
Coordinate nouns
Quantifier (QP; Q1/Q2 not yet implemented)
Number phrases
Specifiers
Adv1
Adj1
N1

The second block of rules repeats until no rule applies. These rules build the consitituents which exhibit free recursion.

26) Cyclic rules: ordered tables of rules covering
Adj2
Adv2
N2
PP (P1/P2 distinction not implemented)
VP
S
Trace insertion
Relative clause
Other clauses [5]

Coordination is attempted after relevant nodes are formed. It assumes that only two nodes of the same type can be coordinated. This scheme naturally runs into the usual difficulties for coordination with a deterministic parser. To try to overcome them, the rules are context sensitive, which provides a form of look-ahead mechanism.

We compared the Left Modifier Grouping (LMG) grammar with a conventional PS grammar which groups right modifiers with the head. In the first case, the grammar could be divided into non-cyclic and cyclic blocks, while in the second, it could not. The first observation is that the division of the grammar in section 3 greatly reduces the number of rules which are considered at any point by the parser. The second is that the Left Modifier Grouping grammar reduces the number of nodes the parser considers earlier than the conventional approach [6]. The combination of these two factors significantly reduces the number of matches the parser attempts. The table in (28) shows data for a comparision in efficiency between two small 9 rule grammars which allow parsing of the sentence (27):

27) A tall man with a hat bought this painting.

28) Comparison of number of rule applications

* LMG grammar for sentence (27)	
* without grammar division	136
* with grammar division	71
*	
* Conventional PS grammar	202

The Left Modifier Grouping approach reduces the number of rule applications required to 35% of those required with a conventional Right Modifier Grouping approach. We conclude that the Left Modifier approach has the double advantage of

1) providing empirically adequate coverage of English in a linguistically principled way, and of
2) increasing the efficiency of the parser

FOOTNOTES

* I would like to thank the following people for valuable discussion and criticism: John Goldsmith, Carmen Lozano, Jonathan Pressler, Richard Culver, Dominique Estival and Tom Reutter.

1. The formulation of the constraint on free recursion is inspired by Emonds' Surface Recursion Constraint.
2. It would also be possible to followed Nanni's analysis more closely and propose that these adjectival phrases are lexicalized.
3. The nature of these nominals can be handle with PS rules, but in the context of our translation system, it is more likely that they will be multiple word lexical entries.
4. The presence of a nominal before the possessive marker makes these sentences much worse, presumably because of the ambiguity generated. Note also that it would be very difficult to parse cases such as these, should one decide to include them in the grammar, because resolution of the ambiguity depends on the semantic content.

5. The grammar we implemented has a third block of rules which deals only with the top-most S. The rules in this last block do not repeat.
6. I have recently learned that parsing grammars have been constructed which use left-associative grammars, and have been found to be extremely efficient (cf. Hausser for an example of such a grammar). Our grammar may differ somewhat, however, in that it is left-associative only in the areas where there is linguistic justification for doing so, as defined by the constraint on free recursion.

REFERENCES

Emonds, J. 1976. A transformational approach to English syntax. Academic Press.
Hausser, R. 1985. Parsing Natural Language using a left associative grammar. Newcat.
Nanni, D. 1980. On the surface syntax of constructions with "easy" type adjectives. Language 56.3/568-581.
Pinkham, J. 1985. The formation of comparative constructions in French and English. Garland Press.